The voice source as a causal/anticausal linear filter

Boris Doval, Christophe d'Alessandro, Nathalie Henrich Bernardoni

The voice source as a causal/anticausal linear filter

Boris Doval, Christophe d'Alessandro & Nathalie Henrich

LIMSI-CNRS
BP133 - Université Paris XI
F-91403 Orsay, France
cda@limsi.fr, doval@limsi.fr

LAM
11, Rue de Lourmel
F-75015 Paris, France
henrich@lam.jussieu.fr

Abstract

A new type of glottal flow model, namely a causal-anticausal linear filter model, is proposed. It is shown that the glottal flow signal can be considered as the impulse response of a linear filter. Then the source/filter speech model can be interpreted as an excitation/filter speech model, the "filter" comprising the glottal flow, vocal tract and radiation components. The spectral features of the voice source models are reviewed, both in the amplitude and phase domains. In the spectral amplitude domain the main features are a spectral maximum (the "glottal formant") and spectral tilt. Evidence for a mixed causal/anticausal phase behavior of the source is given. Then, a causal-anticausal linear filter voice source model is designed. In conclusion, applications of this new approach are discussed for voice quality modification, voice source estimation, voice quality perception.

1. Introduction

The so-called "source/filter" model of speech production, based on the linear acoustic theory, is made of a non-linear volume velocity source, which represents the glottal signal, a time-varying linear filter, associated to the vocal tract, and a radiation component, which relates the volume velocity at the lips to the radiated pressure in the far acoustic field.

For signal processing application, however, another simplification is usually accepted. As the effect of sound radiation seems close to a derivative in frequency domain, the radiation component can be considered as another linear filter component. It is generally a (time-invariant) pre-emphasis filter, i.e. a high-pass first order filter.

In this paper, we shall demonstrate that the voice source component of the source/filter model can also be considered as a linear filter component. This idea is not completely new in the speech processing literature, but we think that the present work will help to establish a link between two approaches of voice source modeling, namely the spectral modeling approach and the time-domain modeling approach, that seemed incompatible, if not incoherent.

On the one hand, in the spectral modeling approach, the source component is considered as a causal linear filter when deriving the linear prediction equations. However, this is lacking rigour in time domain, as the corresponding glottal flow signals do not resemble to any causal impulse response, in the general case.

On the other hand, in the time-domain modeling approach, non-linear models are used (based on polynomial or exponential functions), generally described by model-dependent sets of time-domain parameters. Only partial (and sometimes incompatible) spectral analysis are available.

However, we shall show that both approaches can be envisaged in a unified framework, and that time-domain models can be considered, or at least approximated by a mixed causal-anticausal linear filter.

Then the "source/filter" model can be considered as an "excitation/filter" model. The non-linear part of the source model is associated to excitation (i.e. quasi-periodic impulses), and the mixed causal-anticausal linear part of the model is associated to the filter component, without lack of rigour. The key point being a mixed causal-anticausal filter model, coined herein as the Causal-Anticausal Linear Model or CALM.

The spectral approach to voice source modeling has a number of advantages. Generally, voice quality is better described by spectral parameters, in terms of e.g. spectral tilt, amplitude of the first few harmonics, increase of first formant bandwidth and so one. A second advantage of spectral modeling is signal processing. For voice quality modification, the spectral approach seems much more efficient than the inverse filtering and time-domain modeling approach. A third advantage of spectral modeling is voice source parameter estimation. This point will also be discussed at the end of the paper. Finally, spectral modeling could also be used for voice source parametric synthesis.

The paper is organized as follows. Next section describes the amplitude spectrum of glottal flow models. It is shown that this amplitude spectrum can be described by 2 main components, the so-called glottal formant and a spectral tilt component. Section 3 describes the phase spectrum of glottal flow models. It is shown that a mixed causal-anticausal phase model may fit well to time-domain glottal flow models. Section 4 describes the design of a linear filter glottal flow model. Section 5 discusses of some applications of this new type of glottal flow model, namely voice source modification, voice source estimation and voice source perception.

2. The amplitude spectrum of glottal flow models

2.1. Glottal Formant

It has been shown in a previous paper [11] that the most frequently used glottal flow models (GFM) can be rewritten in terms of the following parameters: E, the value of the glottal flow derivative at the maximum excitation, T_0, the fundamental period, O_q, the open quotient, α_m, the asymmetry coefficient defined as the ratio between the opening duration over the period, and a spectral tilt parameter which can be defined either in the time domain as the time constant T_a of a decaying exponential as in LF or R++ models ([8, 17]) or in the frequency domain as the attenuation TL of a low-pass filter at $3000Hz$ as in Klatt model ([13]). Figure 1 shows one period of a typical glottal flow derivative with those 5 parameters.

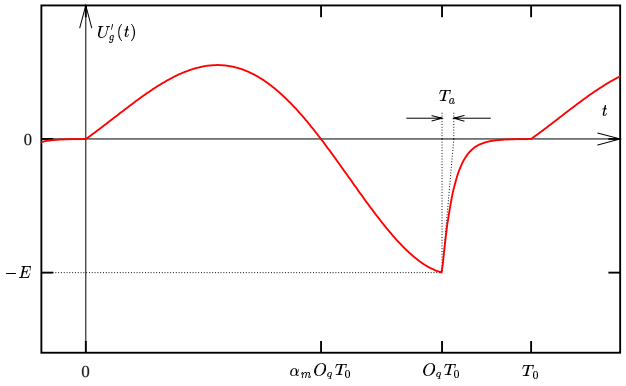

Figure 1: Glottal flow model derivative $U'_g(t)$ and its parameters E, T_0, O_q, α_m and T_a.

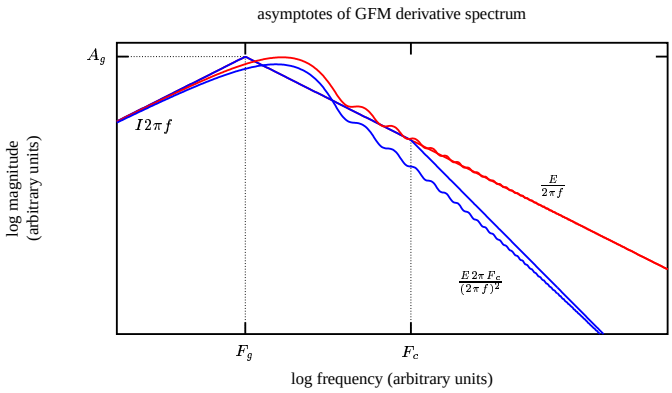

Figure 2: Amplitude spectrum of the glottal flow model derivative. It can be stylized into 3 lines with -6, $+6$ and $-12dB/oct$ slopes.

In the case of abrupt closure (no spectral tilt, i.e. $TL = 0$ or $T_a = 0$), the GFM derivative $U'_g(t)$ can be expressed as:

$$U'_g(t) = E\, n'_g(\frac{t}{O_q T_0}; \alpha_m) * \perp\!\!\!\perp\!\!\!\perp_{T_0}(t) \qquad (1)$$

where $n'_g(t;\alpha_m)$ is a function depending only on the α_m parameter that characterizes a given glottal flow model and $\perp\!\!\!\perp\!\!\!\perp_{T_0}(t)$ is a dirac comb that periodizes the flow.

An analytical expression of the spectrum of glottal flow model derivatives is then obtained and can be studied:

$$\widetilde{U'_g}(f) = E O_q T_0 \widetilde{n'_g}(f O_q T_0; \alpha_m)(F_0 \perp\!\!\!\perp\!\!\!\perp_{F_0}(f)) \qquad (2)$$

where $\widetilde{U'_g}(f)$ and $\widetilde{n'_g}(f;\alpha_m)$ are the Fourier transforms of $U'_g(t)$ and $n'_g(t;\alpha_m)$.

Of great interest is the asymptotic behavior because it shows that the GFM exhibits a maximum in low frequencies, that can be called "glottal formant" (unless it does not correspond to an acoustic resonance). This comes from the fact that, in case of abrupt closure, the asymptotic behaviors near frequency 0 and $+\infty$ are respectively $+6dB/oct$ and $-6dB/oct$:

$$\widetilde{U'_g}(f) \overset{f=0}{\sim} I2\pi f \qquad \widetilde{U'_g}(f) \overset{f\to+\infty}{\sim} \frac{E}{2\pi f} \qquad (3)$$

where I is the total flow over one period (the integral of $U_g(t)$ between 0 and T_0). The crossing point of these 2 asymptotes (see figure 2) is at frequency F_g and amplitude A_g that can be expressed in terms of the time-domain parameters as follows:

$$F_g = \frac{1}{2\pi}\sqrt{\frac{E}{I}} = \frac{f_g(\alpha_m)}{O_q T_0} \qquad (4)$$

$$A_g = \sqrt{EI} = E O_q T_0 a_g(\alpha_m) \qquad (5)$$

where $f_g(\alpha_m)$ and $a_g(\alpha_m)$ are two functions depending only on α_m.

Finally, this asymptotic behavior is the same as that of a second order linear filter, which justifies its being called "glottal formant". For a second order linear filter, besides the position of the asymptotes there is a quality factor that controls the amplitude of the maximum, the spectrum shape being more or less "resonant" (with the same asymptotes). This is also the case for the glottal formant. Figures 3 and 4 illustrate this fact, showing the influence of O_q and α_m on the glottal flow derivative spectrum.

From equation 2 and figures 3 and 4 one can deduce that, for given E and T_0 (as described in [12]):

- The glottal formant frequency is exactly inversely proportional to the open quotient.
- The "quality factor" of the glottal formant is controlled by α_m, the frequency of the glottal formant maximum remaining approximately unchanged.

Notice that a change in T_0 moves the glottal formant frequency in the same direction as O_q does, as can be seen in equation 4 showing that F_g is proportional to F_0. Therefore the glottal formant frequency is not absolute but is linked to the fundamental frequency, and should be expressed as a proportion of F_0 (one should say for instance: "the glottal formant frequency is $1.5 \times F_0$"). Glottal formant frequency values typically range between $0.75 \times F_0$ for O_q values near 1 (soft phonation) and $3 \times F_0$ for O_q values near 0.3 (pressed phonation). The dynamic of glottal formant amplitude values for reasonable values of α_m (between 0.6 and 0.8) and O_q (between 0.3 and 1), E being kept constant, is around $20dB$ but depends on the model ($10dB$ to $22dB$ among models).

2.2. Spectral tilt

In case of smooth closure ($TL \neq 0$ or $T_a \neq 0$), the return phase (or the spectral tilt filter) acts as a first order low-pass filter with a rather high cut-off frequency F_c. Thus, an additional $-6dB/oct$ slope appears after frequency F_c (figure 2):

$$\widetilde{U'_g}(f) \overset{f=+\infty}{\sim} \frac{E 2\pi F_c}{(2\pi f)^2} \qquad (6)$$

2.3. Stylization

Finally, the amplitude spectrum of glottal flow model derivatives can be stylized into 3 lines with $+6dB/oct$, $-6dB/oct$ and $-12dB/oct$ slopes. The first 2 lines build a so-called "glottal formant", a maximum in the spectrum whose frequency (relative to the fundamental) is mainly controlled by O_q and whose amplitude is jointly controlled by O_q and α_m.

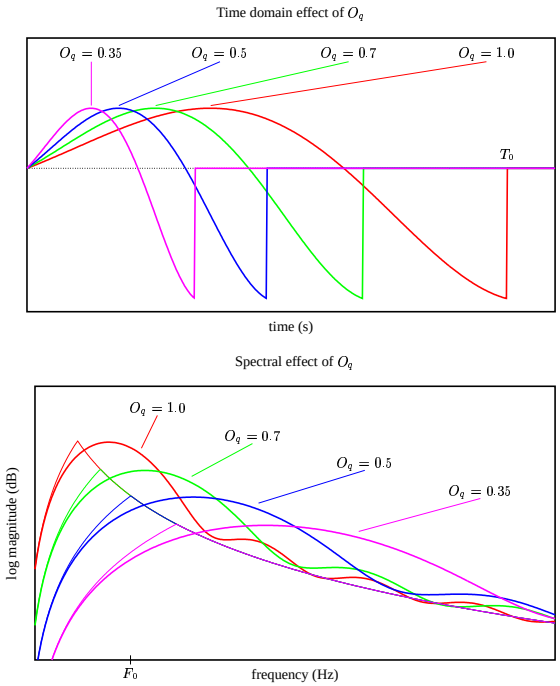

Figure 3: Spectral effect of O_q: the whole spectrum is exactly scaled by the factor $1/O_q$. Notice that the maximum of the spectrum is in the vicinity of the first harmonics.

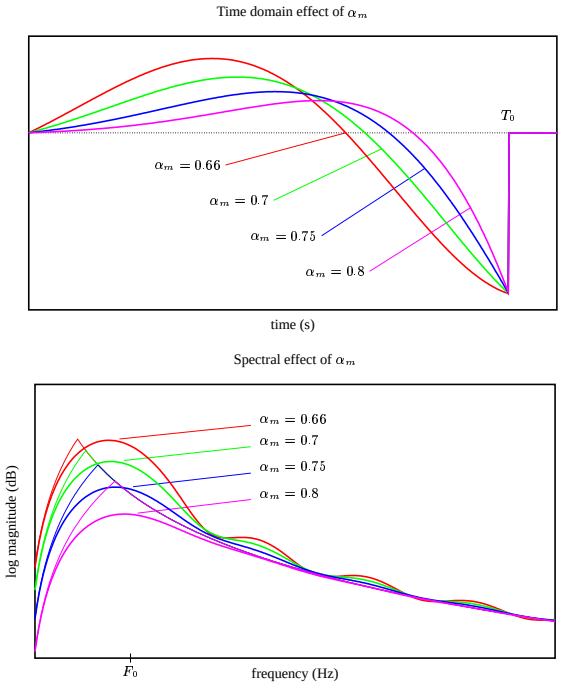

Figure 4: Spectral effect of α_m: the amplitude of the spectrum maximum is reduced/enhanced. This mainly modifies the first harmonics amplitude.

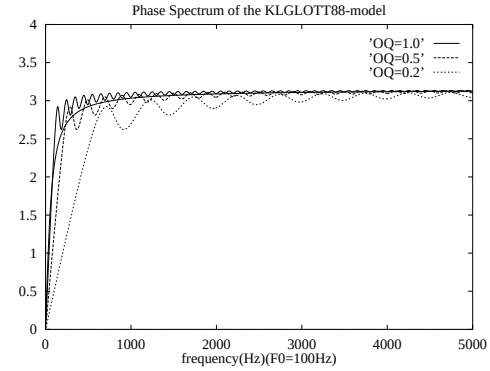

Figure 5: Phase spectrum of the KLGLOTT88 model for various values of O_q. The phase spectrum predicted by an anticausal filter model is also plotted. TL=0

3. The phase spectrum of glottal flow models

3.1. Evidence for anticausality

All glottal flow models (Klatt, Rosenberg, LF, R++, etc.) show an asymmetry that gives more importance to the right part of the flow (figure 6). Most of them have a parameter that can regulate this asymmetry, namely the skewness parameter or asymmetry coefficient.

Evidence for anticausality should be considered as follows: let us extend the glottal flow derivative while keeping its behavior; if we do it to the right (towards positive times) as if it was causal, this will result in an indefinitely increasing (eventually oscillating) waveform; but if we do it to the left (towards negative times) as if it was anticausal, then this will result in a decreasing (eventually oscillating) waveform. Since the behavior has to be stable, the only solution is an extension towards negative times.

This time domain behavior is confirmed by phase spectrum observation where an increasing phase can be observed, at least for low frequencies, as shown on figure 5. This fact has previously been observed and has been used for instance for speech coding as described in [16].

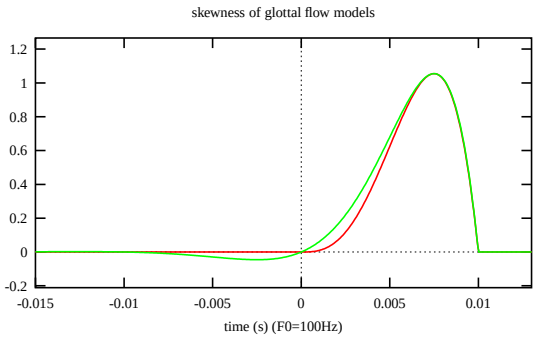

Figure 6: Comparison of the R++ model and the impulse response of the corresponding anticausal filter ($\alpha_m = 0.75$). Notice that both are skewed to the right.

3.2. Anticausality and stability

A linear model of the glottal flow should then have an anticausal behavior. Its impulse response should be zero for times greater than the glottal closure time (at least for the abrupt closure case), but may be non-null for all times before it. Its phase response would show increasing phase, and its transfer function would have a convergence region in the z-plane which is the inside of a circle (this is an equivalent condition to an anticausal impulse response).

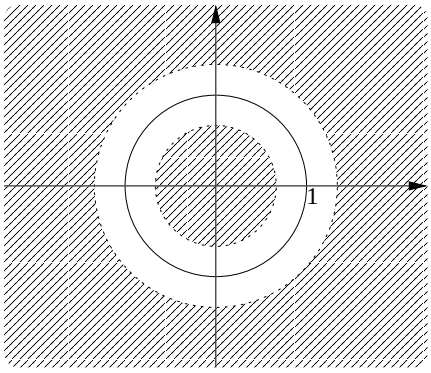

Figure 7: Convergence region for a stable, mixed causal/anticausal filter. The convergence region is in white. Notice that the filter is stable because it contains the unit circle.

But this raises the question of stability. It is often said that a filter that shows poles outside the unit circle is unstable; this is only true within the hypothesis of causality. In fact, a stable causal filter must have its poles inside the unit circle, but a stable anticausal filter must have them outside the unit circle (see [14]). The reason is as follows: the condition for a filter to be stable is that the unit circle must rely in the convergence region. Since for an anticausal filter the convergence region is the inside of a circle, and since it must contain the unit circle for stability, then the poles must be outside the unit circle (a pole cannot be in the convergence region). Figure 7 shows the convergence region for the mixed causal/anticausal case.

3.3. LF model as a truncated linear filter

Observing the analytical expression of LF model, one can see that it is the sum of two parts, namely the open phase $E_1(t)$ and the return phase $E_2(t)$ (the closed phase is zero):

$$\begin{aligned} E_1(t) &= E_0 e^{at} \sin \omega_g t \quad 0 < t < T_e \\ E_2(t) &= -\frac{E_e}{\varepsilon T_a} \left(e^{-\varepsilon(t-T_e)} - e^{-\varepsilon(T_c-T_e)} \right) \quad T_e < t < T_c \end{aligned}$$

where the different variables are explained in [8].

One can recognize that the return phase is the truncated impulse response of a first order continuous time causal filter, and the open phase is the truncated impulse response of a second order continuous time anticausal filter (because a is always positive).

In order to study this anticausal filter, one needs to consider the continuous time transfer function which is obtained by Laplace transform (instead of z-transform for discrete-time filters). In this case, the frequency response relies on the imaginary axis of the s-plane (instead of the unit circle). Then a stable anticausal filter must have all its poles with positive real parts.

Considering the anticausal filter deduced from LF model by extending the open phase towards negative times, one can show that its transfer function (obtained by Laplace transform) is:

$$H_1(s) = -E_0 e^{(a-s)T_e} \frac{(s-a)\sin(\omega_g T_e) + \omega_g \cos(\omega_g T_e)}{(s-(a+j\omega_g))(s-(a-j\omega_g))} \quad (7)$$

where the real part of s must be lower than a. H_1 exhibits 2 poles with positive real parts at $a + j\omega_g$ and its conjugate and one zero at $a - \omega_g cotg(\omega_g T_e)$ (cf figure 8).

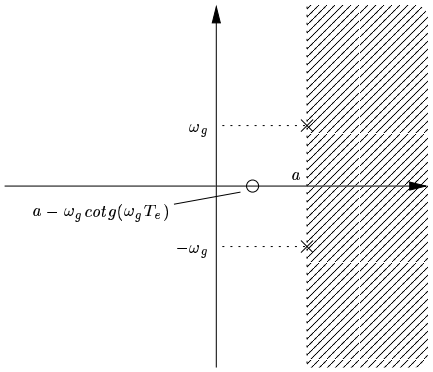

Figure 8: Position of the poles (crosses) and zero (small circle) of the extended version of the open phase of LF model.

One can then consider the LF-model as a left-truncated anticausal impulse response plus a right-truncated causal impulse response.

4. A causal/anticausal linear filter

Pursuing this idea of considering glottal flow models as linear filters, it is tempting to benefit from the large number of results on estimation and modification given by the linear filter theory to inverse the design process, that is to design a filter that could be used for glottal flow modeling (instead of showing that glottal flow models present some of the properties of linear filters).

4.1. Design

To design a digital filter, one has at its disposal many different techniques: either considering the continuous time filter and discretize it with one of the different transformation methods (derivative equivalence, bilinear transform, impulse response discretization, etc.) or designing directly the discrete time filter.

Starting with a continuous filter is interesting because the link between time domain and spectral domain parameters is straight away. Then we model the glottal flow by the impulse response of an all-pole filter that has 2 anticausal poles for the glottal formant and one causal pole for the spectral tilt. To get the glottal flow derivative, one simply has to add a zero at 0. The position of the anticausal poles is defined as follows in function of the time domain parameters :

$$p = a_p +/- jb_p \quad (8)$$

$$a_p = -\frac{\pi}{O_q T_0 \tan(\pi \alpha_m)} \quad (9)$$

$$b_p = \frac{\pi}{O_q T_0} \quad (10)$$

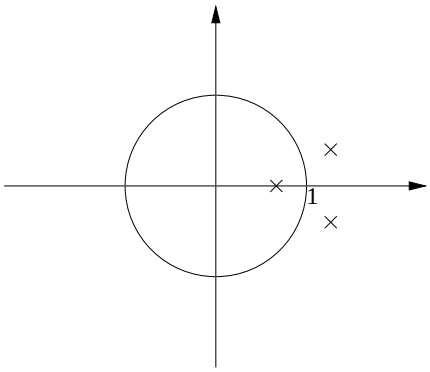

Figure 9: Proposal of a linear causal/anticausal filter to model the glottal flow. The anticausal (outside of the unit circle) pair of poles corresponds to the glottal formant, and the causal (inside of the unit circle) single pole to the spectral tilt.

Among the different transformation methods, the impulse response discretization is a natural way (it corresponds to what is done during digital synthesis) and has the advantage of keeping the same number of poles and not adding any new zeroes, the digital filter still being an all-pole filter. We then obtain a digital filter that has 2 anticausal poles and one causal pole and whose impulse response corresponds to the glottal flow itself (cf figure 9). Its transfer function writes:

$$H(z) = \frac{b_1 z}{1 + a_1 z + a_2 z^2} \tag{11}$$

where the filter coefficients are obtained by (T_e is the sampling rate):

$$a_1 = -2e^{-a_p T_e}\cos(b_p T_e) \tag{12}$$

$$a_2 = e^{-2a_p T_e} \tag{13}$$

$$b_1 = E\frac{\pi^2}{b_p^3} e^{-a_p T_e}\sin(b_p T_e) \tag{14}$$

Then, to get a GFM, one has simply to pass through this filter the signal composed of a dirac at each glottal closure, being cautious that samples are to be computed in the reversed time direction as given by the recurrence equation:

$$y_n = -a_1 y_{n+1} - a_2 y_{n+2} + b_1 x_{n+1} \tag{15}$$

After that, adding a spectral tilt simply involves the causal filter with transfer function:

$$H(z) = \frac{b_{TL}}{1 - a_{TL} z^{-1}} \tag{16}$$

where a_{TL} and b_{TL} are given by:

$$a_{TL} = \nu - \sqrt{\nu^2 - 1} \tag{17}$$

$$b_{TL} = 1 - a_{TL} \tag{18}$$

$$\nu = 1 - \frac{1}{\eta} \tag{19}$$

$$\eta = \frac{\frac{1}{e^{-TL/10.0 * log(10.0)}} - 1}{cos(2\pi\frac{3000}{F_e}) - 1} \tag{20}$$

4.2. Impulse response truncation

These considerations raise the question of the effect of truncation. Truncation in the time domain is equivalent to convolving in the frequency domain with a sine cardinal, thus inducing in the spectrum some regularly spaced zeroes and ripples. This effect can be seen on figure 10.

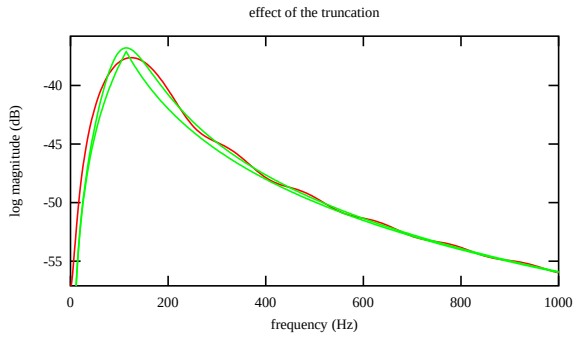

Figure 10: Effect of the truncation on the spectrum. The glottal formant is enlarged and some ripples appear, but the asymptote in $+\infty$ is the same.

5. Discussion

5.1. The CALM and voice quality Perception

The CALM seems to be a good conceptual model for description of voice quality perception. Considering only the quasi-periodic voice source, the CALM parameters can be interpreted almost directly in terms of voice quality.

The most striking voice source parameter (after F_0 of course) is related to spectral tilt. Spectral tilt in the CALM corresponds to the causal part of the model at glottal closure. In terms of spectra, it is related to the number and position of causal poles of the model. In terms of perception, it seems to correspond to the loud/weak voice quality.

Another parameter of interest is the glottal formant position. This parameter is linked to the anticausal part of the model, i.e. the position of anticausal poles. In terms of perception, this seems to correspond to the tense/lax voice quality.

Tenseness and loudness can be varied independently, in the model as well as in actual voice production (with some training however). Of course some more systematic work for relating spectral parameters to voice quality would be needed.

5.2. Voice source parameters estimation

Considering the voice source as a linear filter allows to use techniques derived from linear acoustics to estimate glottal flow parameters. In a recent study [10], we have applied a linear model of the glottal flow in the case of abrupt closure to estimate open quotient by linear predictive analysis of inverse filtered speech. Tested in the case of synthetic vowels, the algorithm performed well for low values of open quotient, but underestimated high values. In the case of a relaxed to pressed phonation, the estimated open quotient values and the measurements done on the corresponding electroglottographic signal were in good agreement.

Yet inverse filtering should be questioned, because it is based on the assumption that the "filter" somehow models the

vocal tract and that the "residual" is linked to the voice source. But if the glottal flow is well represented by a linear filter, this undermines the validity of inverse filtering procedures, because one cannot be sure that some parts of the voice source are not estimated in the "filter" part of the model by the linear prediction procedure. Then, if inverse filtering is often able to give good results, one could think it is due to the fact that LPC is not able to model with equal reliability low frequency and high frequency formants, leaving most of the glottal formant information in the residual.

Since we have shown that the glottal formant can be modeled by an anticausal second order filter, and since the vocal tract must be considered (for physical reasons) to be a causal system, then if we had a way of separating the causal and anticausal parts of the signal, it would allow us to estimate the glottal formant from the signal, even without any inverse filtering procedure. From this "glottal formant signal" the values of O_q, α_m and E_e could theoretically be estimated. Promising results along this line can be found in [5].

5.3. Voice source synthesis and modification

The CALM can be used as an alternative to time-domain models for voice source synthesis. It has the same power as any 5 parameter time domain models like the LF model. Therefore, the same types of sound can be synthesized. The synthesis process consists in designing the appropriate filter as given by equations 8 to 20, and then either convolving the filter impulse response with a train of impulses with fundamental period T_0, or alternatively filtering a spectral comb with fundamental frequency F_0 with the filter amplitude and phase responses.

It must be pointed out that it is also possible to excite the filter with another excitation source, like noise bursts. This may open new ways of synthesising noisy voice qualities.

Spectral processing has long been used for voice quality modification. The CALM provides us with a systematic framework for performing modifications in the time domains, as described in [3]

6. Summary

The time-domain and spectral domain approaches to voice source modeling that have been proposed in the literature seemed somehow incompatible: on the one hand the impulse responses of spectral models used e.g. in linear prediction did not resemble time-domain models, and on the other hand the spectral properties of time-domain models were not recognized as filtering.

We showed in this paper that both approaches can be unified. The price to pay for this common framework is to consider a causal-anticausal linear filter model.

We are convinced that many new analysis, synthesis and processing techniques will take advantage of the CALM, as it open both a unified view of glottal flow models and a better description of the phase of glottal flow signals.

7. References

[1] Alku P. Glottal wave analysis with pitch synchronous iterative adaptative inverse filtering. *Speech Communication 11*, 109–18, 1992.

[2] Alku P., Strik H., and Vilkman E. Parabolic spectral parameter - a new method for quantification of the glottal flow. *Speech Communication 22*, 67–79, 1997.

[3] d'Alessandro C. & Doval B. "Voice quality modification for emotional speech synthesis." Proc. Eurospeech'03, 2003, in press.

[4] Childers D. G., and Lee C. K. Vocal quality factors: Analysis, synthesis, and perception. *J. Acoust. Soc. Am. 90*, 2394–2410, 1991.

[5] Bozkurt B. & Dutoit T. "Mixed-phase speech modeling and formant estimation, using differential phase spectrums." Proc. of Voqual'03 Workshop, Geneva, 2003.

[6] Doval B. & d'Alessandro C. "Spectral correlates of glottal waveform models: an analytic study." Proc. ICASSP'97, 1295–1298, 1997.

[7] Doval B. & d'Alessandro C. "The spectrum of glottal flow models." Notes et document LIMSI, 99–07, 1999.

[8] Fant G., Liljencrants J., and Lin Q. "A four-parameter model of glottal flow." *STL–QPSR*, 85(2):1–13, 1985.

[9] Fant G. "The LF-model revisited. Transformations and frequency domain analysis" STL–QPSR 2–3, 119–56, 1995.

[10] Henrich N., Doval B., and d'Alessandro C. Glottal open quotient estimation using linear prediction. In *Proc. Intern. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications* Firenze, Italy, pp. 12–17, 1999.

[11] Henrich N., d'Alessandro C., Doval B. "Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data" Proc. Eurospeech, Aalborg, Sept. 2001.

[12] Henrich N. "Étude de la source glottique en voix parlée et chantée" Ph.d. thesis, Université Paris 6, France, 2001.

[13] Klatt D. and Klatt L. "Analysis, synthesis, and perception of voice quality variations among female and male talkers." *J. Acoust. Soc. Am.*, 87(2):820–857, 1990.

[14] Orfanidis Sophocles J. "Introduction to signal processing" Prentice Hall International, ISBN 0-13-240334-X, pp. 189–199, 1995.

[15] Rosenberg A. E. "Effect of glottal pulse shape on the quality of natural vowels." *J. Acoust. Soc. Am.*, 49:583–590, 1971.

[16] Sun X.Q., Plante F., Cheetham B.M.G. and Wong W.T.K. "Phase modelling of speech excitation for low bit-rate sinusoidal transform coding" Proc. IEEE Int. Conf. ICASSP'97, Munich, April 1997 (Vol 3, pp. 1691–1694).

[17] Veldhuis R. "A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation" *J. Acoust. Soc. Am.*, 103:566–571, 1998.